

VISUALIZACIÓN DE DATOS: UNA IMAGEN PUEDE VALER MÁS QUE MIL NÚMEROS, PERO NO SIEMPRE MÁS QUE MIL PALABRAS

Data visualization: An image can be worth more than a thousand numbers, but not always more than a thousand words

Alberto Cairo



Alberto Cairo es profesor Knight de la *University of Miami*, y director del programa de visualización del *Center for Computational Science* de la misma institución. Autor de los libros “The functional art: An introduction to information graphics and visualization” y “The truthful art: Data, charts, and maps for communication”.

<http://www.thefunctionalart.com>

<http://orcid.org/0000-0003-4204-0820>

University of Miami, School of Communication
5100 Brunson Drive, Office 5051-L
Coral Gables, Florida 33146, EUA
alberto.cairo@gmail.com

Resumen

En las últimas dos décadas la visualización de datos se ha convertido en una herramienta fundamental para la exploración y comunicación de mensajes complejos en disciplinas como la estadística, la inteligencia de negocios, las ciencias, el periodismo, etc. Este artículo expone las ventajas de la visualización y también algunos de sus principales problemas.

Palabras clave

Visualización; Visualización de datos; Diseño de información; Infografía; Infográficos.

Abstract

In the last two decades, data visualization has become a fundamental tool for the exploration and communication of complex messages in disciplines such as statistics, business intelligence, science, journalism, etc. This article exposes the advantages of visualization and also some of its main problems.

Keywords

Visualization; Data visualization; Information design; Infografía; Infographics.

Cairo, Alberto (2017). “Visualización de datos: una imagen puede valer más que mil números, pero no siempre más que mil palabras”. *El profesional de la información*, v. 26, n. 6, pp. 1025-1028.

<https://doi.org/10.3145/epi.2017.nov.02>

1. Cómo un gráfico puede ayudar a conocer la verdad

El 5 de septiembre de 2017, la publicación ultraderechista estadounidense *Breitbart.com* publicó una noticia con el título “2.139 beneficiarios del programa *DACA* condenados o acusados de crímenes contra americanos”. El programa *DACA* (*Deferred action for childhood arrivals*) protege a aquellos inmigrantes indocumentados que entraron en Es-

tados Unidos irregularmente siendo menores y que hoy son adultos que estudian y trabajan en el país¹.

<http://breitbart.com>

Breitbart es un medio de comunicación abiertamente xenóforo —y, desgraciadamente, leído por millones de personas—, por lo que su objetivo era destacar que (y cito) “el número de [beneficiarios de *DACA*] que son criminales condenados” y que han perdido por ello su estatus protegido es “impresionante”.

Cuando leí esa noticia me dije: El número es “impresionante”, desde luego, ¿pero es impresionante por lo alto o por lo bajo que es? Después de todo, hay más de 800.000 beneficiarios de *DACA* en Estados Unidos.

Una simple calculadora revela que 2.137 es un 0,27% de 800.000. A pesar de que siempre me hayan gustado las matemáticas, tengo una enorme dificultad imaginando la magnitud de cifras con decimales, por lo que he preferido hacer un dibujo. Llamémoslo una “visualización de datos”.

El cerebro humano no está bien preparado para manejar con soltura símbolos arbitrarios y abstractos, como los números

0,27% equivale a 3 personas por cada 1.000 beneficiarios de *DACA*. En la figura 1, cada círculo representa una persona; los círculos rojos son beneficiarios de *DACA* que perdieron sus privilegios por culpa de actividades criminales (figura 1, izquierda).

No parecen tantos, ¿verdad? ¿Y qué pasa si comparamos ese número con la proporción de ciudadanos estadounidenses convictos por una causa criminal en algún momento de sus vidas? El resultado está en la figura 1, derecha.

Así que la cifra que Breitbart puso en el titular de su noticia es “impresionante”, desde luego. Impresionantemente baja. De hecho, por lo menos a primera vista, los beneficia-

rios de *DACA* (3 convictos por cada 1.000) parecen bastante menos peligrosos que los ciudadanos estadounidenses (64 convictos por cada 1.000 personas).

2. Gráficos para descubrir

El famoso estadístico John W. Tukey escribió una vez que el gran valor de las imágenes radica en que muy a menudo nos permiten percibir cosas inesperadas. Tukey se tomó muy en serio sus propias palabras y en los años 70 definió toda una rama de la estadística, el “análisis exploratorio de datos” (*exploratory data analysis*) que, en esencia, consiste en descubrir patrones y tendencias en datos a partir de representaciones visuales como gráficas, mapas, diagramas, etc.

Una visualización de datos no es sólo para ser vista, como si fuese un mero dibujo, sino para ser leída e interpretada con atención

En la actualidad llamamos a estos gráficos visualizaciones de datos. He dedicado más de veinte años de mi vida a estas curiosas herramientas, extensiones de nuestra percepción y cognición.

La visualización de datos está basada en una idea simple, pero también poderosa: el cerebro humano no está bien preparado para manejar con soltura símbolos arbitrarios y abstractos, como los números. Somos capaces de interpretar el sentido de grandes cantidades de cifras sólo indirectamente: por ejemplo, cuando las representamos proporcionalmente por medio de la variación de ciertas propiedades de objetos visuales, como su altura, longitud, tamaño, ángulo, grosor, o color.

Piense en un gráfico de barras: lo que varía en este tipo de gráfico es la altura o la longitud de varios rectángulos, y esta variación debe ser proporcional a los números que representan. En una gráfica de la fiebre de un paciente, por ejemplo, varía la posición de varios puntos y el ángulo y extensión de las líneas que los conectan. En un mapa que represente la tasa de desempleo por provincia puede que varíe la intensidad de color: cuanto menor sea el desempleo, más claro el color; cuanto mayor, más oscuro.

Para comprobar el poder de la visualización de datos para detectar tendencias y patrones, intente extraer alguna historia interesante de la figura 2, una porción de una base

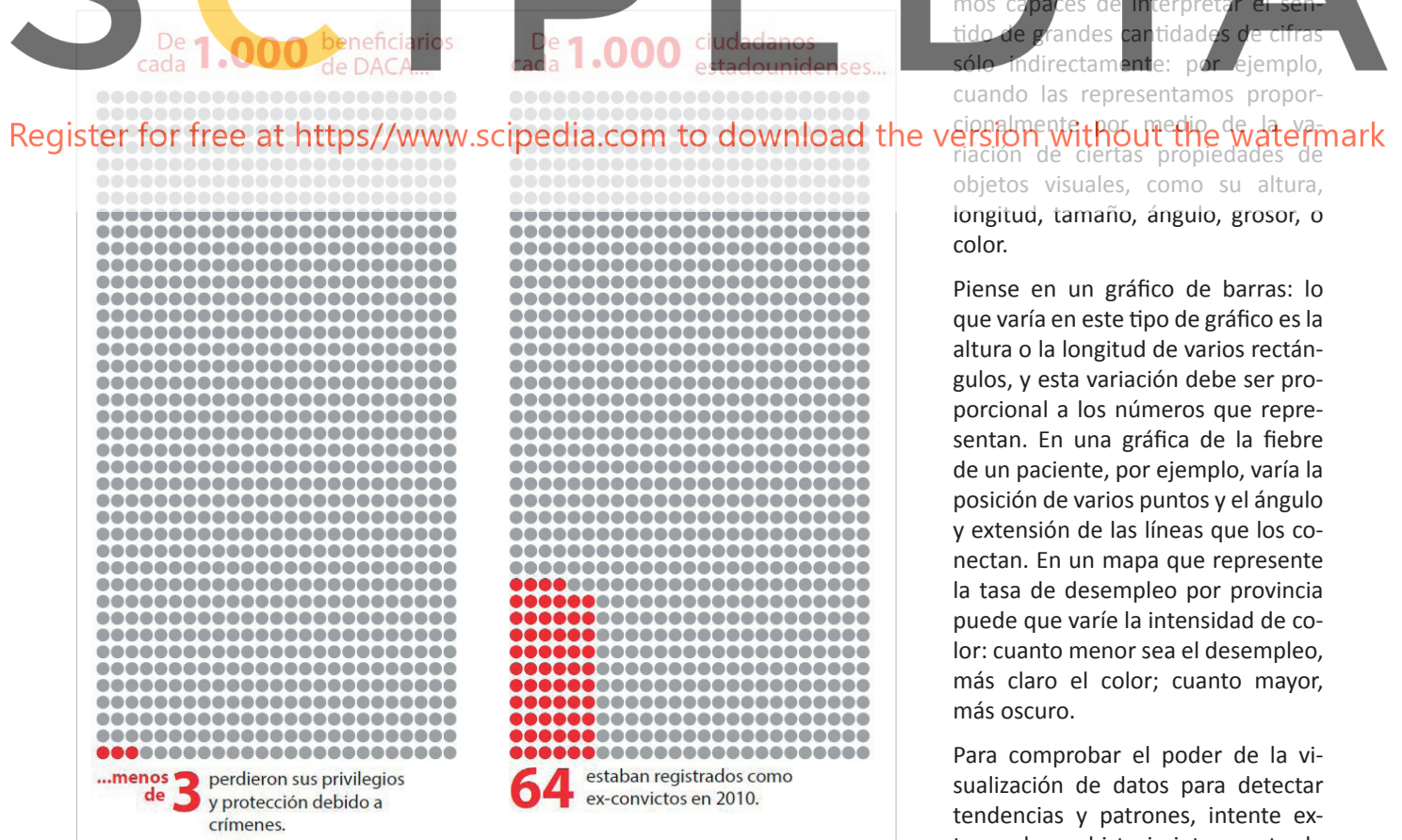


Figura 1. Representación intuitiva de proporciones

de datos de temperaturas globales comparadas con una media de varias décadas del siglo XX (por ello se ven números positivos y negativos: por encima o por debajo de esa media). Estoy seguro de que le costará ver más allá de los propios números.

Veamos ahora qué pasa si, en vez de enseñarle todos los números en una tabla, los transformo en una línea (figura 3). El eje horizontal son los años, desde el 1000 hasta el final del siglo XX. El eje vertical son las temperaturas comparadas con la media de los años 1961-1990 (la línea negra horizontal que recorre el gráfico de parte a parte). La línea azul representa la variación real de las temperaturas en la Tierra, año por año. La línea ondulada negra representa la tendencia general y la franja de color gris claro es el grado de incertidumbre de los datos. La línea roja en la parte superior derecha representa predicciones hechas en el año 1998.

Estoy seguro de que ahora resultará más sencillo interpretar las tendencias en los datos: entre los años 1000 y 1900 (aproximadamente) las temperaturas globales variaron, pero siempre de forma más o menos moderada. A partir del comienzo del siglo XX, por el contrario, las temperaturas se dispararon de forma muy radical, y continúan subiendo hoy en día².

Una imagen —un gráfico, en este caso— vale más que mil números. O más que muchos miles de números. Es por ello por lo que disciplinas tan diversas como la estadística, la inteligencia de negocios, las ciencias y el periodismo han adoptado la visualización como una herramienta de exploración, análisis y comunicación.

Este hecho supone una evolución curiosa. En la cultura occidental, las imágenes siempre han tenido mala prensa. Tradicionalmente se las ha considerado engañosas y poco adecuadas como vehículo de pensamiento, sobre todo en comparación con la lengua hablada y escrita o con la notación matemática. Esto es un error histórico enorme ya que, como espero haber probado, en muchas ocasiones los gráficos son esenciales para la comprensión.

3. Saber interpretar los gráficos

Ahora bien, es cierto que la desconfianza por las imágenes tiene cierto fundamento. Piense en la frase hecha “una imagen vale más que mil palabras”, que transmite la idea de que

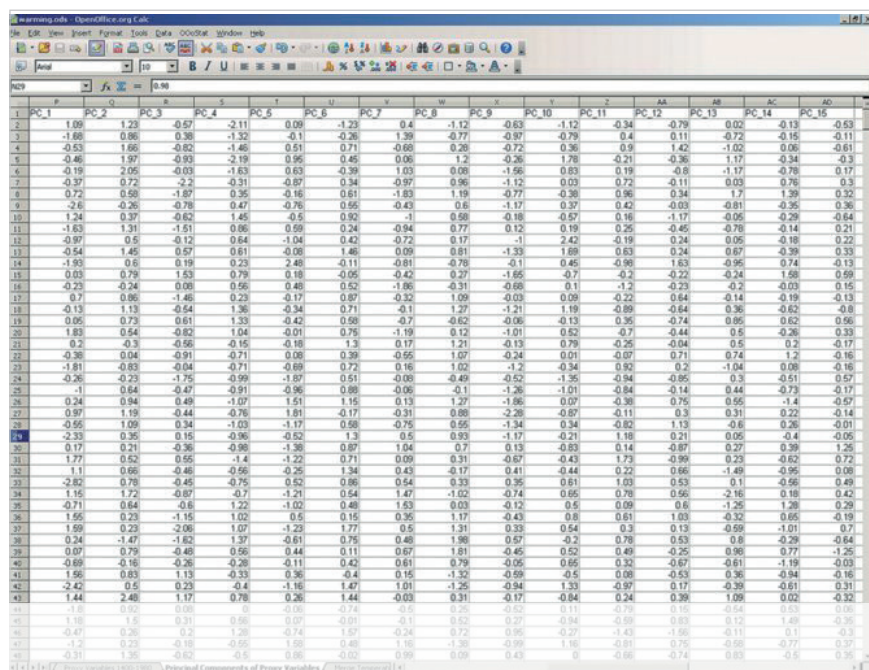


Figura 2. Base de datos de variaciones de temperaturas

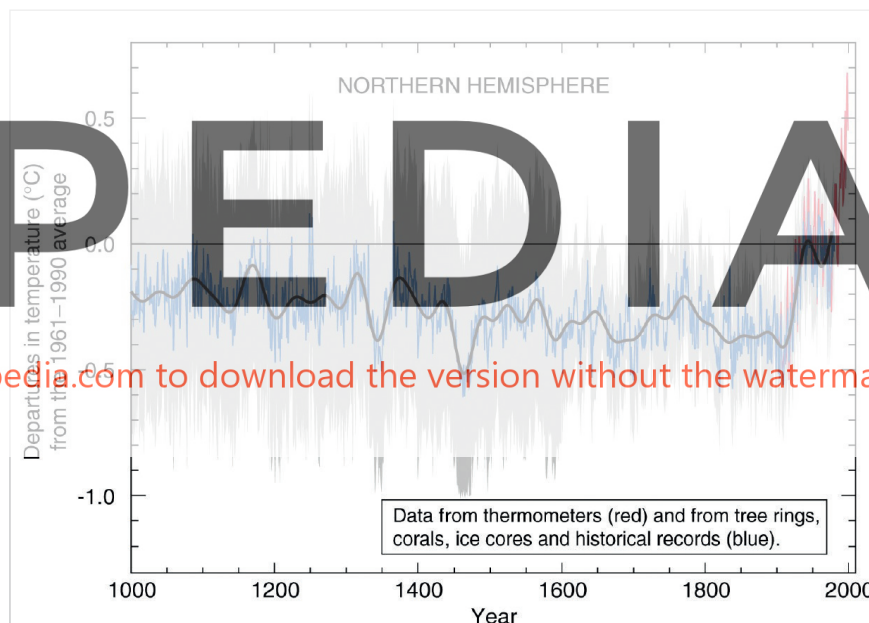


Figura 3. Fluctuación de la temperatura de la Tierra entre los años 1000 y 2000 respecto de la media de los años 1961-1990 (línea negra horizontal)

cualquier imagen puede ser interpretada intuitivamente, de un único vistazo. Esto no es cierto. Una visualización de datos no es algo sólo para ser visto, como si fuese un mero dibujo, sino también para ser leído e interpretado con atención, interrogándose a uno mismo no sólo sobre lo que el gráfico revela, sino sobre lo que puede no estar mostrando.

Véase por ejemplo la figura 4, aparecida en la revista *New England journal of medicine*. Muestra la correlación entre consumo de chocolate per capita en cada país (eje horizontal) y el número de premios Nobel por cada diez millones de habitantes (eje vertical). La correlación es clara: en general, cuanto más chocolate consumen los ciudadanos de un país, más premios Nobel ese país tiene.

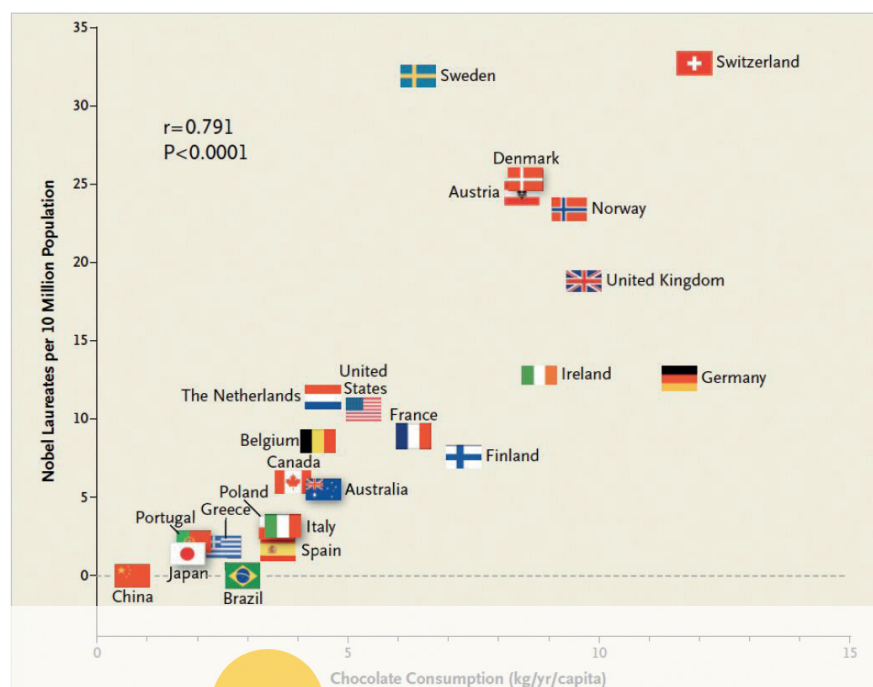


Figura 4. Correlación entre el consumo anual de chocolate per cápita de los países y el número de premios Nobel por cada 10 millones de habitantes. Fuente: Messerli, Franz H. (2012). "Chocolate consumption, cognitive function, and Nobel laureates". *New England journal of medicine*, October 18, n. 367, pp. 1562-1564. <https://doi.org/10.1056/NEJMon1211064>

los ciudadanos de un país, más alimentos suelen consumir y más dinero tienen para invertir en educación —lo que puede conducir a más premios Nobel.

Por ello es importante siempre recordar que puede ser cierto que una imagen o gráfico valga más que mil números, pero no siempre lo es que una imagen valga más que mil palabras. En la mayoría de los casos, palabras y gráficos deben estar unidos para formar argumentos y como sustento de discusiones y debates, reforzándose mutuamente.

Sin palabras que los expliquen, los pongan en contexto, y destaquen sus limitaciones, las visualizaciones de datos a menudo pierden sentido o son fácilmente malinterpretadas. Tales palabras pueden formar parte de los propios gráficos o pueden ser pensamientos que, como lectores, nosotros mismos generemos mientras intentamos descifrarlos.

La próxima vez que vea una gráfica estadística, mapa o diagrama explicativo en un medio de comunicación o en redes sociales, por tanto, no lo comparta después de una mera ojeada. Léalo con atención, piense en la calidad de los datos que muestra y en cómo éstos han sido codificados. Cuéntese a sí mismo, en silencio, lo que percibe. Sea crítico de manera constructiva.

Desde el punto de vista de la simple geometría, este gráfico es correcto. Los números están disponibles en múltiples fuentes y están representados de una forma adecuada: las distancias de cada punto en comparación con los ejes X e Y son proporcionales a los números.

Sin embargo, si uno no tiene cuidado a la hora de leer el gráfico, este puede inducir a error. El coeficiente de correlación entre las dos variables indica también una relación causal: consumir más chocolate puede ayudar a ganar más premios Nobel³.

Enhorabuena. Con esta lectura ha dado su primer paso como visualizador de datos profesional.

Notas

1. <https://goo.gl/7Hc7Su>
2. Este gráfico es tradicionalmente conocido como "el gráfico del palo de hockey" (*hockey-stick chart*) y es obra de varios famosos climatólogos. Se puede leer más sobre él en este artículo: <https://goo.gl/fuSJYe>
3. Esto es exactamente lo que sucedió en este caso, con múltiples medios de comunicación haciéndose eco de esa historia: <https://goo.gl/evWqhF>

Palabras y gráficos deben estar unidos para formar argumentos y como sustento de discusiones y debates, reforzándose mutuamente

El problema, por supuesto, es que tanto la cantidad de premios Nobel como la cantidad de chocolate que un país consume están también relacionadas con muchas otras variables. Entre ellas, la riqueza: Cuanto más adinerados son